DOCUMENT RESUME

ED 458 106 SE 065 241

AUTHOR Tsuei, T. G.; Guo, Sha-Lin; Chen, Jwu E.
TITLE A Practical Approach to the Curriculum of Statistics for Engineering Students.
PUB DATE 2000-08-14
NOTE 6p.; Paper presented at the International Conference on Engineering Education (Taipei, Taiwan, August 13-17, 2000).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Engineering Education; Estimation (Mathematics); Higher Education; *Inferences; *Probability; *Statistics

ABSTRACT

To use the concept of probability and uncertainty estimation is a common practice in daily life. To assist students in learning the fundamentals of probability and statistics, not through the usual class and note taking, a group-learning and practical "statistical system" was designed. By connecting with the main statistical center, students not only can learn the concept of statistics through a group of well-designed activities but also get experience in "crunching" data and "seeing" the results in a more intuitively way to solve their problems. Detailed curriculum contents are presented. The statistical activities consist of two parts. The first part contains games which pertain to the basic understanding of probability, and the other part contains activities which relate to estimation and hypothesis testing of population parameters. (Contains 10 references.) (MM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 458 106

A Practical Approach to the Curriculum of Statistics for Engineering Students

T.G. Tsuei[1], Sha-Lin Guo[2], and Jwu E Chen[3]

[1]*Vice-President office, Ta Hwa Institute of Technology, Hsinchu, Taiwan, ROC, http://www.thit.edu.tw*
Tel:(+886) 3-5927700 ext 2102, Fax:(+886) 3-5926853,adtg@et4.thit.edu.tw
[2]*Ta Hwa Institute of Technology, Hsinchu, Taiwan, ROC, http://www.thit.edu.tw*
[3]*Department of Electrical Engineering, Chung-Hua University, Hsinchu, Taiwan, ROC, http://www.chu.edu.tw*
Tel:(+886) 3-5374281 ext. 8318, Fax:(+886) 3-5374281 ext.8930, jechen@chu.edu.tw

Abstract: To use the concept of probability and uncertainty estimation is a common practice in our daily life. To assist students in learning the fundamentals of probability and statistics, not through the usual class and note taking, a group-learning and practical "statistical system" was setup. By connecting with the main statistical center, students not only can learn the concept of statistics through a group of well-designed activities, but also get experience in "crunching" data and "seeing" the results in a more intuitively way to solve their problems. Detailed curriculum contents are presented. The statistical activities consist two parts. The first are games pertain to the basic understanding of probability, and the other are activities relate to estimation and hypothesis testing of population parameters.

Keywords: statistics, probability, quincunx, estimation, and inference

I. Introduction

It is believed that there is a genuine rule guiding the world. However, in the reality, things are full of "uncertainty" and people simply cannot draw any conclusion with absolute confidence [1-9]. The concept of "probability," therefore, comes into play. Although each "sample" under observation is from the same population, it however behaves with a certain "distribution". In order to get the largest control over the uncertainty, it is essential to know the distribution of samples. In the first part of this paper, we use the ball-dropping experiment to demonstrate the mapping scheme from "path" domain to "result" domain to illustrate the concept of "normal probability distribution."

Statistical inference is another important concept in statistics used when description of the population parameters is required. When limited by time and cost, one may need to do statistical inference based on samples from the population to estimate its corresponding parameters. In the second part of the paper, we present three activities designed with great confidence in estimating population parameters such as the mean and the proportion of the population.

II. Understanding of probability

To illustrate the concept of probability, we have constructed two sets of Quincunxes, which are two re-implementations of Francis Galton's [2] ideas, to demonstrate the rule that governs the distribution of ball-dropping path. The quincunx was a box with a glass face and a funnel inserted over the top. Lead shot of uniform size and weight were released from the funnel to fall through rows of pins. The pins deflected each falling piece of shot to the left or right with equal probability, with each piece of shot finally falling into one of a number of compartments. The dropped shot formed a "normal" curve. Consequently, the second quincunx is based on the idea of modifying the machine so that the shot fell into a series of compartments at intermediate levels. We can find that the piles of shot that fell into each intermediate compartment also formed normal distributions. Further, when the shot in each compartment was released, so as to fall to the bottom of the quincunx, the resulting pile formed a normal curve again. The quincunx provided Galton with proof that normal distribution was normally a mixture of normal distributions. Ideas can also be extended to many other activities, which utilizes different combinations of ball-dropping scheme.

Activity 1: Quincunx-1

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

T. Tsuei

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

By using the output of a designed quincunx machine, the concept of probability distribution can be established. A ball drops from the top of the quincunx machine can go either left or right with equal probability along its path. This means the ball can randomly choose "left" or "right" turn at each level. The "path" of the ball can be defined by a sequence of binary numbers with 0 denote "left" and 1 for "right" turn on each level. We define the group of all possible paths as the "path domain" whereas the ' result domain" is defined as the entire slot numbers where the dropped ball will end up. For example, if there are three levels in a quincunx and the ball always making "left", then the ball will drop in the left most slot on the bottom of the quincunx level, which was labeled as slot "0". The "path" of the ball is thus denoted as (000) corresponding to value 0 in the result domain. When the ball always making "left" except one time, the path of the ball can be one of the following "paths" which are (001), (010), or (100) and the result of the path is 1. This means the ball drops in the second slot from the left which was labeled as slot "1" on the bottom of the quincunx.

Consequently, the "result" of any path can also be defined as the number of making "right" steps or the number of "1"s in the path sequence. Figure 1 shows the mapping from the "path" domain to the "result" domain.

The number of possible path is 2^n where n is the number of quincunx level and each path carries the same probability of $1/2^n$. Thus if there are 3 levels, the probability for result "0" is 1/8 since there is only one possible path (000) map to it. However, the result "1" has probability 3/8 since there are three possible paths (001), (010), and (100) contribute to that result. Consequently the result "2" has probability 3/8 and the result "3" has probability 1/8. This mapping from a value in the "result domain" to its corresponding probability is the so-called Binomial distribution function in Statistics.

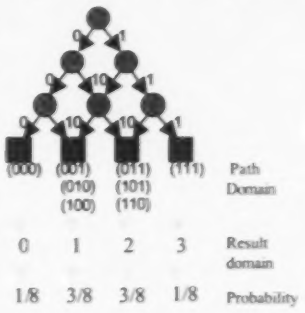

Figure 1. The ball-drop paths.

Activity 2: Quincunx-II

It is to show "a normal mixture of normal distributions is itself normal." From the above analysis, the result domain can be constructed from a sequence of Bernoulli trials [10]. In other words, the distribution of the result domain, which is Binomial, can be expressed as the distribution of sums of independent Bernoulli random variables. Let us assume that a matrix array of pins is used and balls may drop from the top level of the pin array. Because we can see the drop-balls are independent, the distribution of the result domain is just summing up the weighted result of each individual by shifting its position. If the distribution of the weighing on the top level of pin array is binomial, the distribution of the result domain can be easy shown to be binomial. It is noticed that the variation of the new resulting distribution will be larger than the original one.

Activity 3: The error of ball dropping

It is to show how much difference to be observed in an actual ball-drop game by comparing the actual result to the theoretical one. Let students do and record the results by selecting a fixed number of balls. After many trials, let them change the number of balls dropped and repeat the process again. It is to be found that the difference of the mean value will be reduced while increasing the number of balls dropped. The mechanism is the same as that of coin-tossing game governed by the Central Limit Theorem. If we toss a coin 100 times, what is the number of "face" appearance? Moreover, what are the results for 1000 and 10000 times? The result will be close to the half of the number of coins tossed while the coin is uniform. The more times coin-tosses, the more closer to the half.

III. Estimation and hypothesis testing of population parameters

Statistical inference is another important concept in statistics used when description of the population parameters is required. In this section, three activities designed with great confidence in estimating population parameters such as the mean and the proportion of the population are presented.

Activity 4: 95% Confidence interval

Confidence interval is commonly used in estimation. The idea behind the confidence interval comes from the sampling distribution. It was assumed when sampling number is large enough so that the Central Limit Theorem can be applied, then the sampling distribution can be approximated by a Normal distribution. We can construct an interval based on this assumption to contain the population parameter, which we need to estimate with a certain confidence level such as 95%. This means about 95 times the confidence interval we constructed will contain the

population parameter as we do sampling every 100 times. This population parameter can be, for example, the population mean or the population proportion. Figure 2 shows the flow chart of this activity. In this activity, it is to generate a sample from the population using a uniform random number generator.

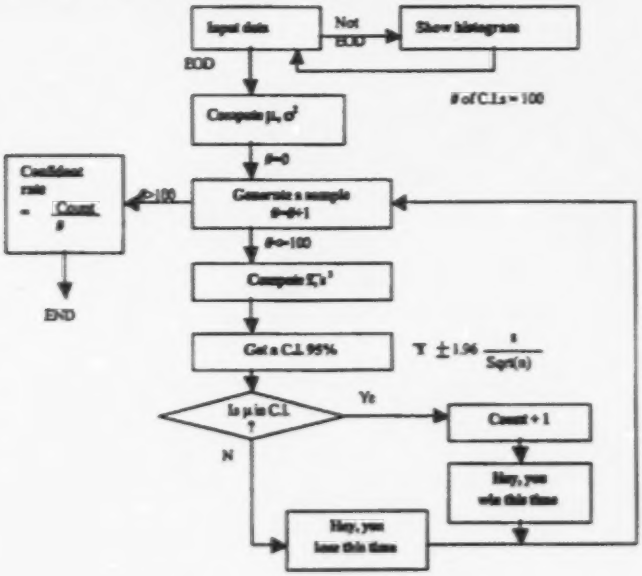

Figure 2. The flow diagram of the sample generation.

Activity 5: Sampling distribution for the sample mean

To estimate the population mean is commonly used in statistical inference. This estimation can be, for example, the average height or weight of a population, the average performance such as the output voltage of a product from a certain procedure. This activity, through the histogram plot for the means of each sampling, illustrates the fact that the dispersion of the sample means is getting smaller when the sampling size is large. This is the one of the applications of the Law of Large Number Theorem and the Central Limit Theorem. Figure 3 shows the flow chart of this activity. It is to note that we will change K's to see if the histogram goes to "normal" when K goes to infinity.

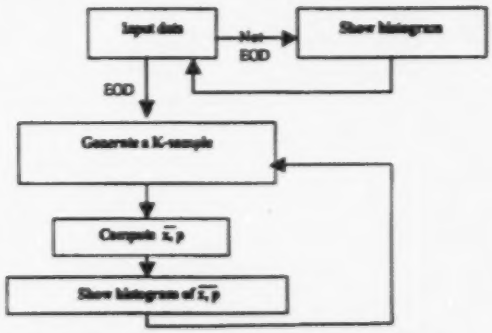

Figure 3. The flow diagram of sampling mean/proportion distribution.

BEST COPY AVAILABLE

Activity 6: Sampling distribution for the sample proportion

This activity is another example of the application of Law of Large Number Theorem and the Central Limit Theorem. It is similar to activity 3 except that the sample mean is substituted by sample proportion. Sample proportion is useful in estimating population proportion, for example, the proportion of possible voters to supporting a certain candidate in election pool or the defect rate of a production line. This activity can demonstrate the fact that the distribution of sample proportion will converge to a normal distribution with small dispersion when sample size is large. It will thus make the estimation more accurate.

IV. Implementations and Experiments

Several prototypes of the system mentioned above have been implemented and tested in the junior course. Figures 4 and 5 show the pin layout of the quincunx-I and quincunx-II implementations. In the quincunx-I, the funnel is fixed and the pin layout is triangle. It allows the balls to drop down from the funnel and bring them back to the top using a motorized belt. It can be used to run any number of balls under user's control. The physical implementation is shown in Figure 6, where the photograph of pin layout is enlarged. The PC through the RS-232 interface controls this machine. Figure 7 shows the control panel in Windows98 platform. The system can be set automatically by user. All relevant information will be gathered, calculated and displayed by control panel. First, the user will determine the number of balls wanted and then the mean and variance can be calculated and modified throughout the process. User can make out their own experiments by using different types, different sizes and different materials of balls to run the games. Will they show the same shape of distribution? It exhibits a "seeing" work. Another seeing work implemented is quincunx-II, as shown in Figure 5, where the pin layout is square and the funnel can be moved left and right. It is used to demonstrate the belief "a normal mixture of normal distributions is itself normal". By the observation, the new distribution is wider than the original one. Any shape of distribution may be adopted as the input pattern. The pin machine can be seen as a scrambler.

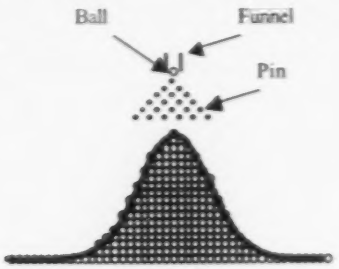

Figure 4. The quincunx-I.

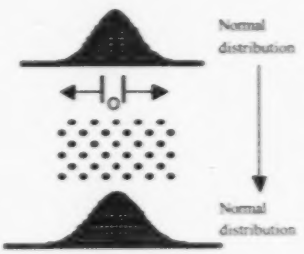

Figure 5. The quincunx-II.

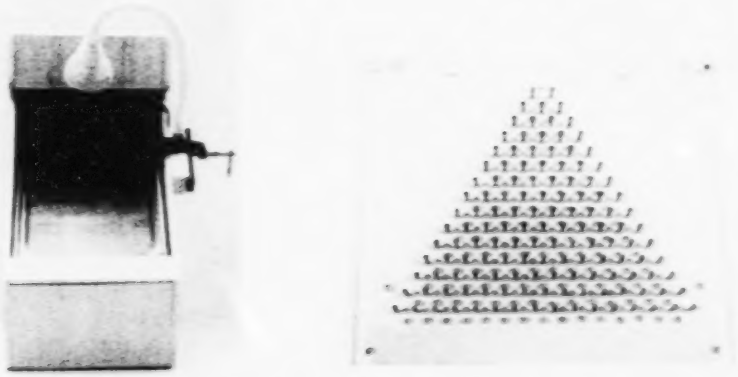

Figure 6. The physical implementation of Quincunx-I and its pin layout.

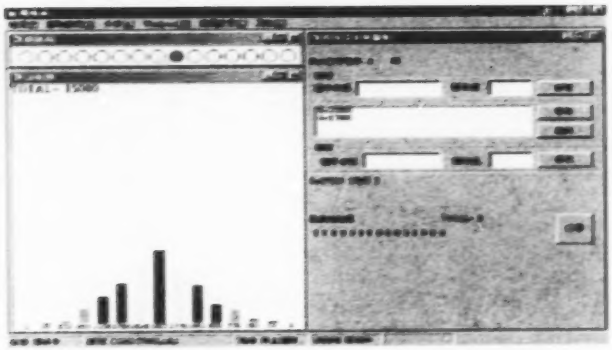

Figure 7. The control panel of quincunx machine in Windows98.

The system of generating random samples has been implemented in the network environment. In a PC classroom, the teacher's computer is perceived as the host and others as terminals. During the activity, the host consequently draws a random number according to the precedent requirement. Then it sends the message to each terminal through the network, calculate and display the results on the screen. Students can easily accept the concept by this way.

V. Conclusions

In the first part of this paper, the ball-dropping experiment *has been shown* to demonstrate the mapping scheme from "path" domain to "result" domain. It also illustrated the concept of "*normal probability distribution*" *and "a normal mixture of normal distributions is itself normal."* Three activities designed with great confidence in estimating population parameters such as the mean and the proportion of the population are also shown in the second part. Most activities have been implemented and tested in the classroom. A quite positive feedback was received from the students.

References

[1] D. J. Bennett, *Randomness*, Harvard University Press, 1998.
[2] Experimentation with the Quincunx, http://www.tsd.sca.edu.au/hist/stats/galton/galton16.htm
[3] THE BEAN MACHINE, http://virtualchaos.com/BMWDisplay.html
[4] The Central Limit Theorem in Action, http://www.rand.org/centers/stat/applets/clt.html
[5] Galton's quincunx," http://stad.dsl.nl/~berric1/.
[6] Normal Approximation to the Binomial Distribution, http://www.ruf.rice.edu/~lane/stat_sim/normal_approx/index.html.
[7] Lightning Calculator Quincunx boards and features, http://www.qualityttg.com/Quincunx.htm
[8] SUMMARISING AND PRESENTING DATA: NORMAL DISTRIBUTIONS, http://www.math.grin.edu/~mooret/main/1-2-5.html.
[9] The Central Limit Theorem and the Quincunx, http://www.users.on.net/zhcchz/java/quincunx/quincunx_8.1.html.
[10] R. V. Hogg and E. A. Tanis, *Probability and Statistical Inference*, Fourth Ed., Macmillan Publishing Company, 1993.

BEST COPY AVAILABLE